Revisiting IP Multicast

By Sylvia Ratnasamy, Andrey Ermolinskiy, Scott Shenker

Presented by Fei Jia

Problems in multicast deployment

- Practical feasibility- Complexity

  - Intra-domain routing: DVMRP, MOSPF, PIM-SM

  - Inter-domain routing: BGMP (Border Gateway Multicast Protocol)

  Overabundance of protocols are hard to ensemble.

- ISPs cannot charge the multicast service efficiently

  - IP multicast is based on an open service model. No mechanism restricts from creating a multicast group, receiving data from a group, or sending data to a group.

FRM (Free Riding Multicast)

- Main idea: An extension to unicast BGP to carry group membership information (that is free riding part)

- Decouple membership discovery from route discovery.
  - Benefits: Once group members are known, any source can construct the multicast tree from its unicast routes

Group membership discovery

- BGP advertisements are augmented with a description of multicast groups

  - Propagated to the global internet

  - Encoded active group addresses

    using bloom filter (which

    produce false positive)

Group Membership Discovery

- What is a group bloom filter?

  After a border router discovers which groups are active in its local domain, it will encode these addresses into a group bloom filter.

- False positive rate

  The filter size L is computed using the false positive rate.

  False positive rate = MIN(1.0, f/(A-G))

  f: an AS is allowed f upstream filters

  A: total size of the multicast address space

  G: the number of groups to be encoded

Multicast forwarding

- At sender domain

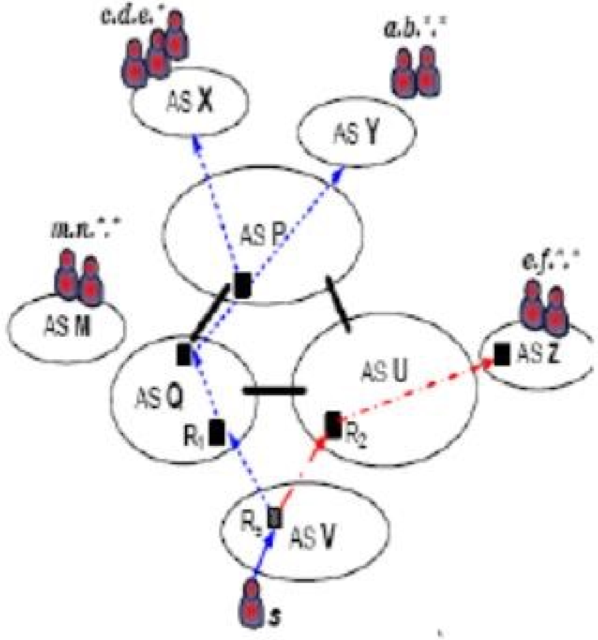
  - Rs (the border router of the sender domain) scans its BGP table to identify the destination prefixes with members in multicast group G

  - Rs constructs the AS-level multicast tree from the individual AS-level paths to each prefix.

  - AS-level tree encoded in shim header and cached

  (Reason:the downstream domains don't know if they are on the path from source to a prefix)

Example



- A packet multicast to G by host s arrives at Rs.
- Rs learns that prefixes a.b.*.*, c.d.e.*, and e.f.*.* have members in G and construct the multicast tree.
- Rs encode 'Q:P','P:X' and 'P:Y' in the packets to R1 and 'U:Z' in those to R2

Multicast forwarding

- At transit domains

  - Routers examine the shim header and forward to neighboring domains based on the tree encoded.

  - The amount of "forwarding" state stored is a list of its neighbor edges without anything else related to the usage of multicast.

Evaluation Assumption

- In a domain of prefix length p, there are $U=2^{32-p}$ users.
- A total of A simultaneously active groups.
- Each user joins k groups selected using some group popularity distribution.

Evaluation-Group Membership

- Memory overhead

  - For storing group membership information

  - Is manageable given current storage cost

  - Example: 1 million simultaneously active groups and 10 groups per user requires nearly 3GB.

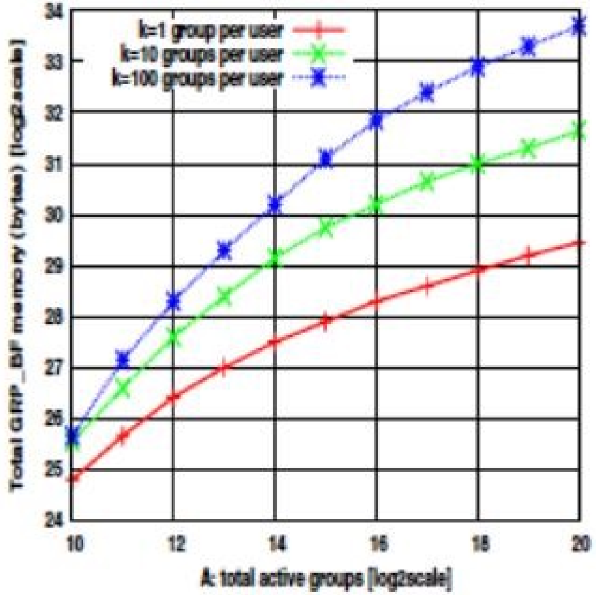
- 

Figure 2: *Total GRP-BF storage per border router.*

Evaluation: Group Membership

- Bandwidth costs

  - For updating group membership

  - A domain updates its membership for group G only when the number of members of G within the domain falls to or rises above, zero.

  - A small fraction of the bandwidth capacity at core BGP routers

Evaluation: Multicast Forwarding

- Bandwidth cost□
  - the per-packet shim header
  - the redundant transmissions required when subrees are large to be encoded in a single shim header

  For 10M users, 99.5% of links see exactly one transmission, and the worst-case per-link-tx of FRM is 157, per-AS is 6950

Group size	Ideal multicast	FRM	per-AS unicast
100	28	28	38
1000	158	159	246
10,000	1000	1012	1962
100,000	4151	4233	9570
1M	8957	9155	21754
10M	15353	15729	39229

Table 2: `total-tx`*: the total number of packet transmissions for increasing group sizes.*

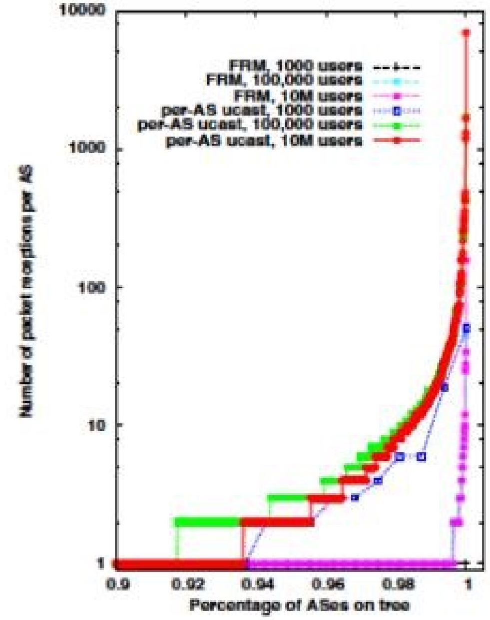


Figure 3: *CDF of* `per-link-tx`*, the transmissions per AS link for FRM and per-AS unicasts.*

Evaluation: Multicast forwarding

- Optimization #1: no leaves
  - Customer Ases at the leaves of the dissemination tree are not encoded into the shim header
- Optimization #2: aggregate links
  - Number of tree edges from an AS A/ A's total edge is large aggregate edges 'A:*'

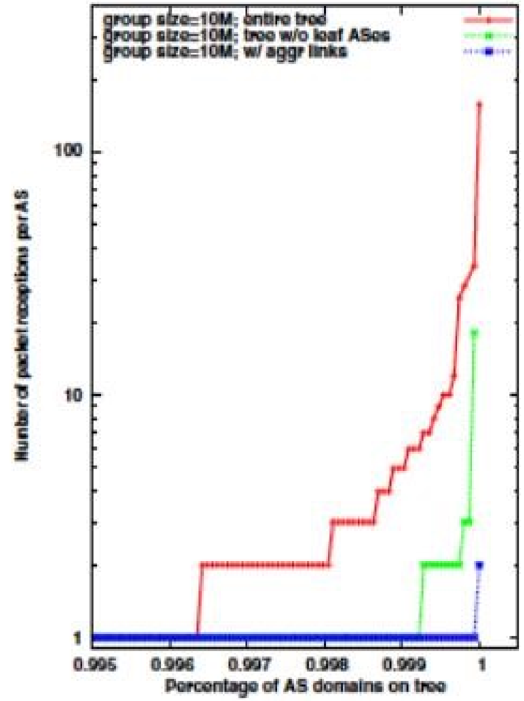


Figure 4: *CDF of transmissions per (AS) link with optimizations to reduce the size of the encoded tree.*

Conclusion

- No distributed tree construction
- Tilts the complexity of route computation to the internals of a router.
- ISP control: ISP can limit service to legitimate users and legitimate groups
- Ease of configuration and allow ISP to work within BGP framework
- More storage and less efficient in bandwidth consumption, but proved reasonable.

Puzzling

- When evaluate forwarding overhead, the paper assumes fixed 100 byte shim header and evaluate the total number of packet transmissions in table 2.
- If compared with total bytes of packet transmissions, what the results will be?

- Any question?